

SYSTEM AND METHOD FOR PERFORMING EFFICIENT DOCUMENT SCORING AND CLUSTERING

Abstract

A system and method for providing efficient document scoring of
5 concepts within a document set is described. A frequency of occurrence of at
least one concept within a document retrieved from the document set is
determined. A concept weight is analyzed reflecting a specificity of meaning for
the at least one concept within the document. A structural weight is analyzed
reflecting a degree of significance based on structural location within the
10 document for the at least one concept. A corpus weight is analyzed inversely
weighing a reference count of occurrences for the at least one concept within the
document. A score associated with the at least one concept is evaluated as a
function of the frequency, concept weight, structural weight, and corpus weight.